

ИСПОЛЬЗОВАНИЕ МОДАЛЬНОЙ ЛОГИКИ ДЛЯ ПРОЕКТИРОВАНИЯ ОБОЛОЧЕК ЭКСПЕРТНЫХ СИСТЕМ

Д.С. Ларионов

Институт "Кибернетический центр" ТПУ

E-mail: dsl@mail.tomsknet.ru

Рассматривается применение модальной логики в качестве механизма вывода экспертных систем. Для примера взята автоэпистемическая логика, являющаяся развитием модальной логики Мак-Дермотта. Описываются такие фундаментальные свойства теории как полнота и адекватность, благодаря которым можно говорить о верных результатах вывода системы. Дается понятие нормальной модальной системы, и указаны аксиомы, за счет которых эти системы обогащаются. Приводится разрешающая процедура для автоэпистемической логики, и описываются ситуации, в которых ее применение будет полезно для механизма вывода экспертных систем.

Введение

В [1] рассмотрен один из способов построения оболочки экспертных систем (ЭС), где механизм вывода системы основан на логическом выводе, опирающемся на логику предикатов первого порядка. Данный вывод ЭС корректен, т.е. используемая аксиоматическая система вывода

$$A \vdash A \quad (A1),$$

$$A, A \supset B \vdash B \quad (MP)$$

адекватна и полна. Однако корректность системы (A1), (MP) сохраняется лишь при так называемом «четком» выводе, где есть лишь два понятия: «истина» и «ложь». Реальные ЭС имеют дело с неполной информацией, поэтому данная система использует коэффициенты определенности (КО) относительно фактов, с которыми она работает. Так как КО вводятся эмпирически (отчасти на основе теории условной вероятности), то корректность системы (A1), (MP) ставится под сомнение в силу отсутствия четкого обосновывающего математического аппарата для КО. Как вариант, сохраняющий с одной стороны математическую строгость, с другой — разноплановую оценку фактов системы, можно рассмотреть модальную логику.

Нормальные модальные системы и логика Мак-Дермотта

Модальная логика использует операторы \Box (общность) и \Diamond (существование), которые можно рассматривать в качестве аналогов КО при использовании логики предикатов. Согласно [2] в основе нормальной модальной системы лежит четверка, состоящая из:

- множества всех теорем логики высказываний, область действия которых распространена на формулы модального языка высказываний L ;
- схемы аксиомы дистрибутивности

$$\Box(p \supset q) \supset (\Box p \supset \Box q), \quad (K);$$
- правила modus ponens

$$\frac{(p, p \supset q)}{q},$$
- модального правила вывода необходимости

$$\frac{p}{\Box p}.$$

Обогащать нормальную модальную систему можно за счет аксиом:

$$- \Box p \supset p \text{ (знания)} \quad (T);$$

$$- \Box p \supset \Box \Box p \text{ (позитивной интроспекции)} \quad (4);$$

$$- \Diamond p \supset \Box \Diamond p \text{ (негативной интроспекции)} \quad (5).$$

Выбор модальной системы зависит от моделируемого понятия. Система, обладающая всеми указанными аксиомами, называется $S5$ (или $KT45$). Так же следует отметить слабую $S5$ -систему (или $K45$).

Мы рассматриваем так называемые *немонотонные логики*. Как известно, в формальной системе дедукции классической логики отношение выводимости \vdash обладает свойством *монотонности*: если $\{p_1, \dots, p_n\} \vdash q$, то $\{p_1, \dots, p_n, r\} \vdash q$. Другими словами, в монотонной системе полученный результат не опровергается дальнейшими, т.е. множество посылок можно увеличить лишь вместе с увеличением множества заключений. Немонотонная система не обязана (и, как показывает практика, ни в коем случае не должна) обладать этим свойством. Из нее можно получить, используя правила вывода, различные несовместные множества формул. Отдельное взятое множество заключений существенно зависит от порядка применения правил вывода. Чтобы избежать наличия нескольких несовместных множеств возможных заключений, обычно ослабляют монотонность до *ограниченной монотонности*: если $\{p_1, \dots, p_n\} \vdash r$ и $\{p_1, \dots, p_n\} \vdash q$, то $\{p_1, \dots, p_n, r\} \vdash q$. Ограниченная монотонность не позволяет моделировать взаимно несовместимые альтернативы рассуждений.

В классической логике теорема — это тавтология и результат общезначимых выводов. В немонотонной логике за «теоремы» следует принять нечто другое.

Если немонотонная система должна *сама обеспечивать* отказ от своих выводов, то ее правила вывода должны быть *модифицируемы*. Их применение может динамически блокироваться. Для этого они сами оснащаются условиями применения, проверка которых динамически изменяется вместе с множеством посылок. Это правила *немонотонного вывода*. Их условия (называемые *предусловиями*) по-

звоняют до вывода проверять выполнимость некоего утверждения вместе с уже выведенными утверждениями в этой системе из существующего множества посылок. Однако есть опасность закливания в определении и интерпретации отношения выводимости, т.к. для проверки невыводимости некоего утверждения немодифицируемые правила должны апеллировать к содержащей их системе вывода.

Мак-Дермотт и Дойл [2–4] предложили изящный метод, позволяющий избежать закливания при задании правил немонотонного вывода. Они ввели неконструктивную характеристику устойчивых множеств взаимно выполнимых формул, немонотонно выводимых из некоего набора посылок. Эти множества суть решения некоторого уравнения, являющиеся неподвижными точками и связанные с отношением выводимости, определяемым данной немонотонной системой. Соответствующая система может рассматриваться как классическая модальная аксиоматическая система, пополненная правилом вывода *выполнимых* утверждений.

Автоэпистемическая логика и ее свойства

В качестве основополагающей логики автор предлагает взять *автоэпистемическую логику*, являющейся реконструкцией (т.е. модификацией с дополнениями) немонотонной логики Мак-Дермотта и имеющей предметом своей формализации *интроспективные и идеально разумные* рассуждения об исходном множестве предположений. *Интроспективные рассуждения* зависят от текущего состояния знаний субъекта; эти рассуждения модифицируемы, т.к. зависят от изменчивого состояния знаний. Под *идеально-разумными* понимаются рассуждения, идеализированные в 2-х аспектах: можно выводить только ожидаемые логические следствия из исходного множества предположений, и все эти логические следствия надо принять во внимание. Идеально-разумные рассуждения позволяют осуществить формализацию выражений вида «если я не предполагаю, что p подтверждается, то подтверждается q ». Основная ценность немонотонной логики Мак-Дермотта заключена в методе неподвижной точки, используемом для характеристики устойчивых множеств заключений немонотонной системы, а также в применении модальной логики для формализации модифицируемых рассуждений. Примером автоэпистемической логики может служить *логика веры и знания*, позволяющая определить и аксиоматизировать различные эпистемические понятия (предположение, знание, обоснованное знание). Здесь формула $\Box p$ интерпретируется: «предполагается, что p подтверждается». Подобные рассуждения немонотонны, т.к. множество основных предположений субъекта может со временем меняться, что чревато противоречиями для некоторых выводов. Такими интроспективными рассуждениями можно моделировать

многочисленные виды модифицируемых рассуждений. Автоэпистемическую логику можно рассматривать как результат реконструкции немонотонной логики Мак-Дермотта, состоящей в замене парадигм *выводимости* и *выполнимости* на формализацию *интроспективных способностей рассуждений*. К тому же, автоэпистемическая логика и логики умолчаний имеют одинаковую формальную выразительность, хотя их области применения могут быть различными.

Теорией называют множество формул модального языка L_p , автоэпистемической теорией – подмножество T из L_p , представляющее какое-то *полное* и *легальное* множество предположений, которое *идеально разумный* субъект может построить на основе множества A исходных предположений.

Автоэпистемическая теория T семантически *полна* тогда и только тогда, когда она содержит все формулы, подтверждающиеся во всех автоэпистемических моделях T . (Интуитивно: T полна тогда и только тогда, когда T содержит все формулы, которые данному субъекту семантически позволено вывести в предположении истинности всех его гипотез).

Автоэпистемическая теория T семантически *легальна* относительно множества A основных предположений тогда и только тогда, когда любая автоэпистемическая интерпретация T , являющаяся моделью A , является также и моделью теории T .

Для уточнения семантических свойств T необходимы следующие определения:

1. *Интерпретация высказываний* автоэпистемической теории T приписывает значение истинности формулам множества T . Приписывание подчиняется классическим правилам оценки сложных формул логики высказываний. Оно придает произвольное значение истинности пропозициональным константам и формулам вида $\Box p$ («предполагается p »).
2. *Модель высказываний* автоэпистемической теории T – это интерпретация высказываний из T , в которой подтверждаются все формулы из T .
3. *Автоэпистемическая интерпретация* автоэпистемической теории T – это интерпретация высказываний из T , для которой всякая формула вида $\Box p$ подтверждается тогда и только тогда, когда p принадлежит T . Таким образом, при автоэпистемической интерпретации формула $\Box p$ («предполагается p ») подтверждается в том и только том случае, если p принадлежит множеству T нашего субъекта.
4. *Автоэпистемическая модель* автоэпистемической теории T – это автоэпистемическая интерпретация, в которой подтверждается всякая формула из T .

Важнейшие свойства автоэпистемической логики строятся вокруг следующего понятия: говорят, что автоэпистемическая теория T *устойчива*, если T есть множество формул из модального языка L_p , удовлетворяющее следующим условиям:

1. Если $\{p_1, \dots, p_n\} \subseteq T$ и $\{p_1, \dots, p_n\} \vdash q$, то $q \in T$.
2. Если $p \in T$, то $\Box p \in T$.
3. Если $p \notin T$, то $\neg \Box p \in T$.

Первое правило утверждает, что мыслящий субъект предполагает все логические следствия из уже предполагаемого (это обязательно, если требуем идеальной разумности субъекта). Второе правило гарантирует, что формула «предполагается p » принадлежит множеству T предположений этого субъекта, если p – предложение этого субъекта. Последнее правило утверждает, что формула « p не предполагается» фигурирует во множестве T предположений того же субъекта, если формулы p там нет.

Устойчивое расширение T определяется так: T является устойчивым расширением A тогда и только тогда, когда T есть множество тавтологических следствий из множества $A \cup \{\Box p \mid p \in T\} \cup \{\neg \Box p \mid p \notin T\}$.

Автоэпистемическая теория T называется *теорией*, основанной на множестве дополнительных аксиом A , если все формулы из T фигурируют среди тавтологических следствий из множества $A \cup \{\Box p \mid p \in T\} \cup \{\neg \Box p \mid p \notin T\}$.

Тогда верны следующие утверждения (важные для применения в механизме вывода ЭС):

- Автоэпистемическая теория T семантически *полна* тогда и только тогда, когда она *устойчива*.
- Автоэпистемическая теория T семантически *легальна* (в смысле выполнима) относительно множества A основных предположений тогда и только тогда, когда она *основана* на A .

Построенная Мак-Дермоттом теория немонотонного вывода обнаруживает некоторые странные особенности, а именно: каждая теорема из *немонотонной* $S5$ -системы является теоремой из *монотонной* системы $S5$. Этот важный теоретический факт можно объяснить, привлекая автоэпистемическую логику. Мак-Дермотт рассматривает неподвижные точки T системы вывода своей немонотонной логики, приложенной к множеству A дополнительных аксиом. Эти точки можно сравнить с максимальными легальными множествами предположений идеального разумного субъекта. *Немонотонная* $S5$ -система содержит схему *аксиомы знания*, т.е. $\Box p \supset p$. Для автоэпистемического анализа немонотонности эта схема кажется слишком обременительной, т.к. она утверждает: «все, что предполагается, истинно». Это приемлемо в *логике знания*, но не в *логике веры*, однако при построении немонотонной системы использование логики знания неподходяще. Схему *аксиомы знания* приходится исключить. Получаем *слабую* $S5$ -систему (обозначается $K45$), система вывода которой дает всегда одни и те же устойчивые расширения. Таким образом, автоэпистемическая логика может обойтись без явного упоминания схем модальных аксиом.

Устойчивое расширение как основа разрешающей процедуры

Описанная выше семантика автоэпистемической логики удобна тем, что с ее помощью можно характеризовать предположения субъекта, не зависимо от того, разумен он или нет. Однако для разумного субъекта она неконструктивна в том смысле, что в ней нет правил, позволяющих оценивать предположения субъекта о сложных формулах, исходя из его предположений о составных частях формул и связях между ними. С этой целью Мур [5] предложил альтернативную семантическую характеристику своей автоэпистемической логики. Эта новая семантика, основанная на понятии *возможных миров*, позволяет построить конечные модели для автоэпистемических теорий. Основным результатом, на котором базируется эта характеристика, является следующее утверждение:

- T есть множество формул, подтверждающихся во всех мирах $S5$ -полной структуры тогда и только тогда, когда T является устойчивой автоэпистемической теорией. (T1)

Таким образом, любая автоэпистемическая интерпретация L устойчивой автоэпистемической теории T может характеризоваться структурой K типа $S5$ и некоторой оценкой V . Структура возможных миров специфицирует предположения идеального разумного субъекта, тогда как оценка определяет то, что действительно подтверждается в реальном мире. Точнее, *автоэпистемическая интерпретация* L автоэпистемической теории T есть пара $L = (K, V)$, где

- K – $S5$ -полная структура (представленная множеством своих возможных миров, каждый из которых символизирован множеством подтверждающихся позитивных и негативных пропозициональных констант),
- V оценка истинности в реальном мире для пропозициональных констант из L_p .

Автоэпистемическая теория T является множеством всех формул, истинных во всех мирах из K .

L называется *автоэпистемической моделью* теории T , составленной из множества формул, истинных во всех мирах из K , когда любая формула из T подтверждается в L .

С помощью следующего результата можно проверить, является ли автоэпистемическая интерпретация L для T автоэпистемической моделью для T .

- Если $L = (K, V)$ – автоэпистемическая интерпретация для T , то L является автоэпистемической моделью для T тогда и только тогда, когда V – элемент из K , что означает выполнимость оценки V вместе с данной оценкой, задаваемой в одном из возможных миров структуры K . (T2)

В силу теоремы (T1) автоэпистемическая теория T устойчива, если ее можно представить $S5$ -полной структурой возможных миров. Чтобы авто-

эпистемическая теория T была *устойчивым расширением* надо еще, чтобы T была основана на множестве A исходных предположений. Иначе говоря, любая автоэпистемическая модель посылок A должна также быть и моделью для T . В силу теоремы ($T2$) оценка реального мира каждой из этих автоэпистемических моделей посылок должна быть выполнимой вместе с оценкой в одном из возможных миров структуры K для автоэпистемической интерпретации.

Возможны различные разрешающие процедуры для автоэпистемической логики. Автор предлагает следующую разрешающую процедуру для автоэпистемической логики: пусть A – множество посылок.

1. Строим все оценки, возможные для пропозициональных констант, появляющихся в A . Они будут характеризовать $S5$ -полные структуры возможных миров языка для A .
2. Выбираем структуры возможных миров, для которых любая формула из A подтверждается во всех мирах.
3. Для каждой из этих структур K строим все автоэпистемические интерпретации (K, V) , соответствующие всем оценкам V пропозициональных констант, которые появляются в A .
4. Проверяем для каждой автоэпистемической интерпретации (K, V) выполнение или невыполнение утверждения «любая формула из A истинна в (K, V) тогда и только тогда, когда V принадлежит K ». В случае выполнения K характеризует некое устойчивое расширение для A .
5. Чтобы проверить принадлежность данной формулы устойчивому расширению, представленному посредством K , выясняем, подтверждается ли эта формула в K .

Пример использования разрешающей процедуры

Пусть множество исходных предположений $A = \{\neg \Box p \supset q\}$. Оно содержит только одну формулу, означающую «если я не предполагаю p , то q подтверждается». Докажем, что:

- Идеально разумный субъект, имеющий множество исходных предположений A , состоящее из одного элемента, получит устойчивое расширение T , содержащее высказывание q , но не p .
- Не может существовать ни устойчивого расширения, содержащего p , но не q , ни устойчивого расширения, содержащего одновременно p и q .

Предположим, что какое-то устойчивое расширение T для A содержит q , но не p . В этом случае $S5$ -полной структурой, связанной с множеством предположений T , будет $K = \{\{q, p\}, \{q, \neg p\}\}$. Эта структура отражает два возможных мира: в первом подтверждаются q и p , во втором – q и $\neg p$. Истинные формулы во всех мирах этой структуры соответствуют формулам, предполагаемым данным субъектом. Таким образом, T – устойчивая автоэпистемическая теория.

Рассмотрим все автоэпистемические интерпретации для T . Они составлены из структуры возможных миров, отражающих предположения данного субъекта, и из произвольной оценки V истинности в реальном мире. В рассматриваемом языке лишь две пропозициональные константы; значит, оценок – четыре: $\{p, q\}$, $\{p, \neg q\}$, $\{\neg p, q\}$, $\{\neg p, \neg q\}$.

Из четырех автоэпистемических интерпретаций $L = \{K, V\}$ только первая и третья превращают в истинное основное предложение $\neg \Box p \supset q$. Т.к. оценка V для каждой из этих интерпретаций совпадает с оценкой возможного мира данной структуры, то теория T основана на множестве посылок A , и, следовательно, является устойчивым расширением A .

Предположим, что T содержит p , но не q . Структура возможных миров будет тогда $K = \{\{q, p\}, \{q, \neg p\}\}$. Рассмотрим оценку $V = \{\neg p, q\}$. Она служит основой автоэпистемической интерпретации $L = \{K, V\}$ теории T , подтверждающей A . Так как V не соответствует оценке какого бы то ни было мира этой структуры, то T не может быть устойчивым расширением A .

Предположим, что T содержит p и q . Единственно возможным является мир $\{p, q\}$. Т.к. существует автоэпистемическая интерпретация, подтверждающая A и заданная этим возможным миром и оценкой $V = \{\neg p, \neg q\}$, для которой V не соответствует единственно возможному миру данной структуры, то T не может быть устойчивым расширением A .

Заключение

Использование автоэпистемической логики в качестве механизма вывода ЭС имеет положительную сторону в том, что она позволяет характеризовать заключения, ожидаемые от системы, способной к совершенной интроспекции. Эта способность важна для ЭС в том смысле, что облегчается процесс построения механизма объяснения для вывода фактов и обоснования решений при условии постоянно меняющегося состояния знаний субъекта. Корректность результатов вывода таких систем гарантирована благодаря четкому математическому обоснованию полноты устойчивых автоэпистемических теорий.

Следует отметить, ссылаясь на [6] и [7], что наличие эвристик является неотъемлемой частью реально работающей ЭС, поэтому попытка «зажать» механизм вывода только в жесткий математический аппарат не принесет практического успеха. Следует применить более гибкий подход с использованием аппарата семантических и нейронных сетей, т.к. для использования только автоэпистемической логики нужны неограниченные временные и компьютерные ресурсы (в силу идеализации рассуждений) и требуется эвристика для ограничения этих рассуждений.

В [1] указано, что оболочка ЭС с прямым и обратным механизмом вывода на основе ($A1$), (MP) реализована на языке логического программирова-

ния Visual Prolog, основанного на работе процедуры доказательства метода линейной резолюции [8] с использованием хорнова дизъюнкта. Предлагаемый аппарат модальной логики с некоторыми

ограничениями и допущениями также можно реализовать на языке Visual Prolog, обладающим специальными средствами доказательства, опирающимися на логические рассуждения [9].

СПИСОК ЛИТЕРАТУРЫ

1. Ларионов Д.С., Новосельцев В.Б. Оболочка экспертной системы на основе нечетких рассуждений с прямым и обратным выводом // Математическое моделирование. — 2002. — Т. 14. — № 9. — С. 48–52.
2. Логический подход к искусственному интеллекту: от классической логики к логическому программированию / Пер. с франц. А. Тейз, П. Грибомон, Ж. Луи и др. — М.: Мир, 1990. — 432 с.: ил.
3. McDermott D., Doyle J. Non-Monotonic logic 1 // Artificial Intelligence. — 1980. — V. 13. — № 1–2. — P. 41–72.
4. McDermott D. Non-monotonic logic 2: non-monotonic modal theories // J. ACM. — 1982. — V. 29. — № 1. — P. 34–57.
5. Moore R.C. Possible-world semantics for auto-epistemic logic // Proc. AAAI—Workshop on Non-Monotonic Reasoning. — October 1984. — New Paltz, N.Y., 1984. — P. 344–354.
6. Jackson P. Introduction to Expert Systems. — Addison Wesley Longman Limited, 1999. — 542 p.
7. Уотермен Д. Руководство по экспертным системам. Пер. с англ. — М.: Мир, 1989. — 388 с.: ил.
8. Стерлинг Л., Шапиро Э. Искусство программирования на языке Пролог. Пер. с франц. — М.: Мир, 1990. — 235 с.: ил.
9. Люггер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е изд.: Пер. с англ. — М.: Вильямс, 2003. — 864 с.: ил.